Case Studies of Value at Risk vs. Expected Shortfall

Niall Whelan

April 24, 2024

The concept of Expected Shortfall (ES) is commonly used as a risk measure and undergirds the new Fundamental Review of the Trading Book (FRTB) capital requirements for traded products of banks. However, it can seem a bit abstruse so here we discuss three specific examples which can be helpful for explaining and understanding its use in a financial context.

1. a single credit-risky bond: demonstrates the improved sensitivity of ES

2. two credit-risky bonds: demonstrates better reflection of risk diversification

3. log-normal distribution: demonstrates improved behaviour for large equity volatility or term

The idea of Expected Shortfall (ES) as a "coherent" risk measure in finance dates to 1999 with an article by Artzner et al. They introduced it to overcome deficiencies of the more traditional Value-at-Risk (VaR) risk measure. For normal distributions, the two measures behave well and indeed are proportionally related. Therefore, standard examples showing the deficiency of VaR typically consider distributions with "lumpy" outcomes such as jumps and discontinuities. Here we point out that even the standard log-normal distribution of stock returns provides an example where VaR can behave perversely.

To set terms, VaR is a percentile. It is the value of some random variable (such as a future stock price or market risk loss) you would not expect to exceed more than a specified fraction of the time. ES is the expected[1] outcome of some random variable (such as a future stock price or market risk losses), conditional on its exceeding a specified amount.

Case 1 – A single bond

Consider a portfolio consisting of a single zero-coupon corporate bond with a maturity of one year. If our risk horizon is one year, then there are only two outcomes: either the bond pays back the principal or it does not, due to default. We can think of this random variable as having a Bernoulli distribution.

[1] "Expected" in the statistical sense of the term, which can be understood as the average.

Imagine that at a moment in time we assess the one-year probability of default (PD) to be 70bps. We can represent the distribution graphically as shown in Figure 1.

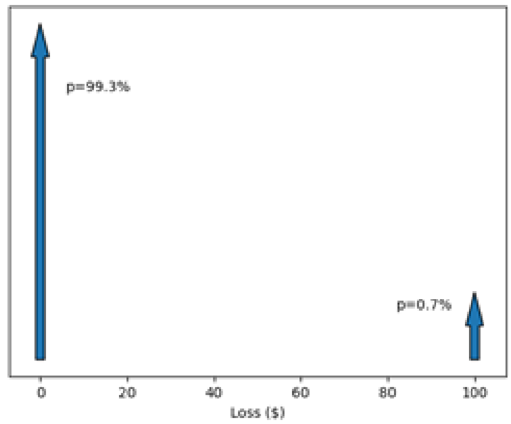

Figure 1: Loss distribution for a credit risky bond

The locations of the arrows indicate the amount of loss ($0 or $100) and the heights represent their probabilities (they are not to scale.) If we are interested in VaR at the 99^{th} percentile, then we conceptually break the arrow at zero loss into two components as shown in Figure 2.

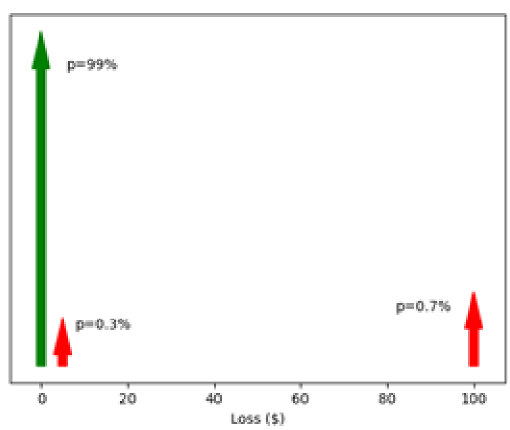

Figure 2: Loss distribution for a credit risky bond, refined

We have now colored the arrows and split the single arrow at Loss=0 into two. We have shifted one arrow to the right graphically but they both represent the same zero loss. We can then think of the small red arrow at zero together with the red arrow at loss=100 as representing the 1% tail. The green arrow at

zero represents the 99% of the loss distribution not in the tail.

Returning to the definition of VaR, we note that it is the point at the threshold of the 99% loss. In our picture this is the boundary between the green arrow and the smaller red arrow. Since both of those arrows are at Loss=0, we conclude that the VaR is 0. Which seems to imply no risk in the position.

This, however, is unsatisfactory since there clearly is some risk. Furthermore, imagine that credit markets take a negative turn such that a week later we reassess PD to be 120bps. At that point VaR would have jumped to the full principal amount (assuming zero recovery.) So at some point our measure of risk would have discontinuously jumped as the assessed probability of default passed through 100bps. These twin defects of being zero when there is non-zero risk and having discontinuities as market conditions change are inadequacies of VaR.

PD (bps)	VaR($)	ES($)
70	0	70
120	100	120

Table 1: Summary of discussion for a single bond (assuming $100 notional.

Expected Shortfall on the other hand behaves coherently. In the example above, we simply calculate the weighted average loss of the two red arrows[2] so that

$$ES = \frac{0.3\% * \$0 + 0.7\% * \$100}{1\%} = \$70. \tag{1}$$

In other words, for a bond with a 70bps probability of default and using the same threshold as the VaR measure, the ES would be 70% of the notional amount. As credit markets worsened, ES would increase continuously up to the point that default probability becomes 100bps (i.e., the specified percentile of 1%), after which it caps out at 100% of the notional amount.

We also plot this graphically for a range of PD values in Figure 3.

Case 2 – A pair of bonds

Another problem with VaR, which is addressed by using ES, is how it performs under diversification. We expect that diversifying a portfolio should lessen the risk and this is indeed satisfied by ES but not by VaR. To see this in an example, we add a second bond to our portfolio. We assume that bond A is the same as in Example 1, i.e., has a PD of 70 bps and zero recovery. We add a second bond B, which has a PD of 50bps and a loss given default of 70% (i.e., a recovery of 30%.) Both bonds have a maturity of one year.

We consider three portfolios: P_A consists only of bond A, P_B consists only of bond B, and P_{AB} consists of both bonds A and B. For each of P_A and P_B, we use our assessment in Example 1 to see that since the probability of loss is less than 1%, each has a VaR measure of zero. Next, we consider P_{AB}: there

[2]The weighting is taken care of by dividing the sum by 1%, whose presence in the denominator is due to the definition of ES as a conditional expectation.

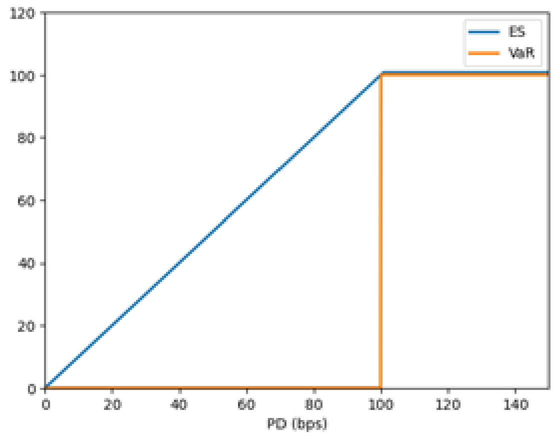

Figure 3: ES and VaR as a function of PD

are four possible outcomes as tabulated in Table 2 where we assume the bond defaults are independent so we can multiply probabilities.

Scenario #	Description	Probability	Cumulative Probability	Probability Exceeding 99%	Loss ($)
1	Neither defaults	98.8035%	98.8035%	NA	0
2	Only B defaults	0.4965%	99.3%	0.3%	70
3	Only A defaults	0.6965%	99.9965%	0.6965%	100
4	Both default	0.0035%	100%	0.0035%	170

Table 2: Four possible outcomes for P_{AB}, ordered by the loss experienced. The "Probability Exceeding 99%" column indicates the relative contribution of each scenario to the 1% loss tail.

To determine the VaR, we look at the ordered loss events and assess where the cumulative loss probability first exceeds 99%, since that will correspond to the 1% worst loss outcome. That is clearly Scenario #2. (As in Example 1, we can conceptually break scenario 2 into two arrows with identical losses and probabilities of 0.1965% and 0.3% and recognise that the 99% threshold is between those two arrows.) We conclude that for the combined portfolio the VaR is the loss associated with scenario 2, i.e. $70.

This is a perverse outcome since it says that by putting the two bonds into a single portfolio, the risk measure (a loss of $70) is worse than if we had considered each bond separately (a loss of $0). This runs counter to Markowitz's famous statement that "Diversification is the only free lunch in investing."

Now we consider the corresponding ES for the three portfolios. For P_A the ES is $70, as we discussed above. Analogous reasoning for P_B is $35 (Of the worst 1% of outcomes, half the time there is zero loss and half the time there is a loss of $70, so the conditional expectation is $35.) Then for P_{AB}, we

Portfolio	VaR($)	ES($)
P_A	0	70
P_B	0	35
$P_A + P_B$	0	105
P_{AB}	70	91.25

Table 3: VaR and ES for the credit risky bond portfolios.

consider the cases in the Table 2. We calculate the weighted average loss of the three scenarios which lie in the 1% tail, using the "Probability Exceeding 99%" column as weights to arrive at:

$$ES = \frac{0.3\% * \$0 + 0.6965\% * \$100 + 0.0035\% * \$170}{1\%} = \$91.25. \quad (2)$$

The results of this discussion are tabulated in Table 3. The row called $P_A + P_B$ is simply the straight sum of the rows above and is provided for reference. Our expectation is that a reasonable measure of risk should register less risk for the combined portfolio P_{AB} than for the sum of the risks for each portfolio separately as given in the row called $P_A + P_B$. That is clearly true for ES but not for VaR.

Case 3 – Lognormal distribution

It is natural to attribute the perversities of VaR discussed in Cases 1 and 2 to their lumpy, discontinuous natures. There are discrete outcomes and as the probabilities for those outcomes cross the percentile threshold, VaR behaves discontinuously. It is tempting then to think that VaR will behave reasonably for smooth distributions. We show that intuition is not always true, using an example of a distribution of immediate relevance to finance: the lognormal distribution describing equity prices under the standard Black-Scholes framework.

We start though with a quick discussion of VaR and ES for the standard normal distribution. If there is a variable x, with zero mean and unit standard deviation σ, (*i.e.*, $x \sim \mathrm{N}(0,1)$ its percentile is

$$\mathrm{VaR}_p = \alpha, \quad (3)$$

where α is the number of standard deviations which correspond to the specified percentile for a standard normal (*i.e.*, $\alpha = \mathrm{N}^{-1}(p)$) and p is the specified percentile. For example, $\alpha = 1.282$ for $p = 90\%$.

The expected shortfall for a specific percentile is available in closed form. By definition we need to calculate the expected value of x, conditional on its exceeding VaR. So, we need to do this integral

$$\mathrm{ES}_p = \frac{1}{\sqrt{2\pi}} \frac{1}{1-p} \int_{\alpha}^{\infty} dx x e^{-x^2/2}. \quad (4)$$

The first prefactor is the normalisation of the pdf and the second is so that we appropriately condition on being in the tail (the same reason we divided by 1%

VaR metrics		ES metrics	
p	α	p'	α'
0.990	2.326	0.974	1.943
0.977	2.000	0.942	1.572
0.900	1.282	0.754	0.688

Table 4: Values of p and p' which lead to the same numerical values of VaR and ES for a normal distribution. For convenience we also show the corresponding values of α and α'.

in the ES examples in Eq.1 and Eq.2.) The integral can be expressed in closed form as

$$\mathrm{ES}_p = \frac{1}{\sqrt{2\pi}} \frac{e^{-\alpha^2/2}}{1-p}. \tag{5}$$

By definition, ES will exceed VaR so a straight comparison of their values may not be all that useful. The mere fact of switching from VaR to ES means that the numerical value of the risk measure will increase. While true, this is not particularly interesting in and of itself. Rather, we are more interested in the qualitative behaviour of these measures as key features of the model change, as we observed in Cases 1 and 2. Therefore, in going from using VaR to using ES, it is common to used redefined values p' and α' so as to have the same numerical value of VaR and ES for a normal distribution. This is the tack taken by the Basel Committee on Banking Supervision as part of their Fundamental Review of the Trading Book where they switched from VaR at 99% to ES at 97.5%.

We can do this through a simple numerical "goal seek" where we equate VaR_p with $\mathrm{ES}_{p'}$. In Table 4 we tabulate a few examples of equivalent values. We now discuss the lognormal distribution. Here we will assume that we are interested in the value of a stock in one year's time. We also make the simplifying assumptions that we expect zero growth so that the expected value of the stock is its current value, and that the stock's current value is unity. This simplifies the analysis, but we will relax those assumptions at the end. We also say that the annualised volatility of the stock is σ. In that case the logarithm of the stock price is a normal distribution with mean $-\sigma^2/2$ and standard deviation σ. In symbols we express this as

$$\mathrm{X} = \log(S) \sim \mathrm{N}\left(-\sigma^2/2, \sigma\right). \tag{6}$$

The non-zero mean of the distribution of X ensures that the expected value $\mathbb{E}[S] = 1$, where S represents the random value of the stock in a year's time and X is its logarithm. The percentile for X is then $\alpha\sigma - \sigma^2/2$, which is simply the equation for the VaR of a standard normal but scaled by the volatility and shifted by the mean. This means that the VaR of the stock price S is the exponential of the percentile of X

$$\mathrm{VaR}_p = \exp\left(\alpha\sigma - \sigma^2/2\right). \tag{7}$$

As we increase σ, we are adding to risk and we expect VaR to increase. This is true up until $\sigma = \alpha$ after which VaR starts to decrease. We will explain the reason for this behaviour later but for the moment it is enough to flag that this is a problem. We expect that increasing volatility should increase VaR, and yet it does not.

We now consider what happens to the Expected Shortfall. We note that

$$\begin{aligned}
\mathrm{ES}_p &= \mathbb{E}[S\,|S > \mathrm{VaR}_p] \\
&= \mathbb{E}[\mathrm{VaR}_p|S > \mathrm{VaR}_p] + \mathbb{E}[S - \mathrm{VaR}_p|S > \mathrm{VaR}_p] \\
&= \mathrm{VaR}_p + \frac{\mathbb{E}[(S - \mathrm{VaR}_p)^+]}{1-p}.
\end{aligned} \tag{8}$$

In the second line we have simply added and subtracted the conditional expectation of VaR_p. In the third line we have made the following manipulations to the two terms, respectively:

1. the conditional expectation of a scalar is just that scalar
2. we have removed the explicit conditioning in favour of defining the expectation in terms of the positive value of the argument. We have also had to explicitly insert the normalisation in the denominator to account for the conditional expectation, similar to its presence in (Eq. 4)

This is a nice result since it explicitly shows that ES is equal to VaR plus an additional positive amount. We can go one step further and notice that the expectation in the second term is nothing but the value of a call option with initial price of one, zero risk-free rate, term of one year, volatility of σ, and strike of VaR_p. We denote the value of that call as $C(\mathrm{VaR}_p, \sigma)$, as the usual Black-Scholes expression but where we suppress the dependence on spot, risk-free rate and term since they are all fixed in this example. We conclude

$$\mathrm{ES}_p = \mathrm{VaR}_p + \frac{C\,(\mathrm{VaR}_p, \sigma)}{1-p}. \tag{9}$$

A final point is that in comparing VaR with ES, we will use different percentile thresholds as discussed in the discussion of the normal distribution. In Figure 4 we plot VaR_p using $p = 0.9$ and $\mathrm{ES}_{p'}$ using $p' = 0.754$, per Table 4. We plot the two measures as a function of σ.

We see that for small values of volatility the two measures are almost identical. This is because in that limit the lognormal distribution and the normal distribution are very close. And we have selected p and p' such that the measures are identical for a normal distribution, so the two measures should be close for small vol. However, as we increase the volatility, the ES is consistently larger and, as noted, the VaR reaches a maximum and starts to decrease. ES, on the other hand, increases monotonically as we increase volatility, which is a consistent behaviour for a risk measure.

This is really what we wanted to show; that even for a smooth distribution the VaR measure can have perverse behaviour. And that ES behaves more intuitively in the same situation.

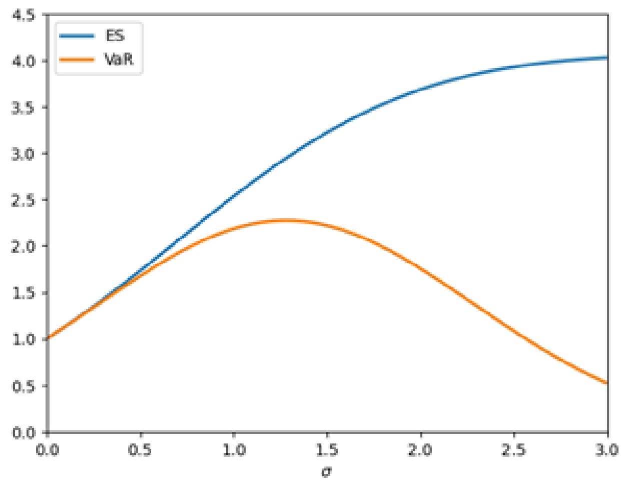

Figure 4: ES and VaR as a function of volatility for a log-normal distribution.

We close with a few comments:

- we can generalise to terms different from one year by interpreting the horizontal axis to be in units of $\sigma\sqrt{T}$. This means that even for moderate volatility, we would observe VaR decreasing so long as the term was sufficiently large
- we can generalise to initial values of share price not equal to unity and risk-free rates different from zero by interpreting the vertical axis to be in units of $S_0 e^{rT}$
- VaR approaches zero with large σ, since we are constraining the distribution to have unit mean even as the volatility gets large. The distribution manages this by putting more and more of its mass close to zero even as its tail gets wider. This means that for any specified percentile, eventually that point on the distribution will be part of the mass of the distribution close to zero, despite the fact that the expectation is unity. This can be understood as a breakdown of ergodicity, a theme that has been explored by Ole Peters[3]
- ES approaches $1/(1-p)$ in the same limit of large σ. The defining equation of ES involves the expectation of S, conditional on S being larger than VaR. But in the large σ limit most of the expectation comes from large values of S, due to the fat tail. So, the conditional expectation of S approaches the unconditional expectation of S, which is unity, other than the normalisation constant of $1/(1-p)$ arising from the conditioning

[3] *The ergodicity problem in economics* https://www.nature.com/articles/s41567-019-0732-0